International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN(P): 2319-3972; ISSN(E): 2319-3980
Vol. 5, Issue 1, Dec – Jan 2016; 25-34

International Academy of Science,
Engineering and Technology
Connecting Researchers; Nurturing Innovations

GENETIC DIVERSITY ANALYSIS IN INDIAN MUSTARD [*BRASSICA JUNCEA* (L) CZERN & COSS]

URMIL VERMA[1], N. K. THAKRAL[2] & NEERU[3]

[1]Department of Mathematics, Statistics and Physics, India

[2,3]Oilseeds Section, Department of Genetics and Plant Breeding; CCS Haryana Agricultural University, Hisar, India

ABSTRACT

Genetic diversity plays an important role in plant breeding. Identification of diverse parents in any crop species like Indian mustard is the pre-requisite. Selection, which is the basis of every breeding programme operates only on variation which is of genetic nature and a wide range of variability present in any crop always provides the better chances of selecting the desirable types. The emphasis of this study was to study the genetic divergence in Indian mustard and grouping them into different clusters based on yield and yield contributing traits for the hybridization programme. Principal Component analysis (PCA) revealed that the first seven PCs explained about 74 % of the total variation and thus indicating that the traits viz., leaf width, leaf length, days to maturity, days to 50% flowering, no. of siliquae on main shoot, siliqua density, seed yield/plant, oil content, main shoot length, 1000-seed wt. and siliqua length are more useful i.e. the higher loading displaying variables. Genetic divergence analysis was performed on the basis of Discriminant analysis using Mahalanobis' D^2-statistic. Based on the relative magnitude of D^2-values ; 60 genotypes of Indian mustard were grouped into five clusters and plant height, no. of siliquae on main shoot and days to maturity were found the best discriminatory characters for the selection of diverse genotypes.

KEYWORDS: Genetic Diversity, Principal Component, Cluster Mean, D^2-Value, INTER-Cluster Distances

INTRODUCTION

In a crop improvement programme, the measurements are taken on several characters because of their inter-relationships; however, a breeder may be interested in selecting only few important characters in which the improvement is needed. Selection, which is the basis of every breeding programme, operates only on variation which is of genetic nature and a wide range of variability present in any crop always provides the better chances of selecting the desirable types. A number of statistical procedures have been proposed from time to time for selection of important characters. Step-wise regression analysis and Principal component analysis (PCA) can be used by researchers for the purpose. PCA has an edge over the other as it removes multicollinearity among the independent variables. Genetic diversity plays an important role in plant breeding because hybrids between lines of diverse origin generally display a great heterosis than those between closely related strains. Multivariate analysis is a useful tool in quantifying the degree of divergence between biological population at genotypic level and to assess the relative contribution of different components to the total divergence. (Zahan *et al.* 2008). Several measures are being used to assess the genetic diversity among plant populations. Of these measures, multivariate analysis {Fisher (1936), Jolliffe (1972), Johnson and Wichern (2006) etc.} provides the most reliable information. Among the multivariate procedures, Mahalanobis (1936) generalized distance (D^2) has been used extensively.

Indian mustard [*Brassica juncea* (L) Czern & Coss.], the crop considered for this study is one of the most important oilseed crops of the country occupying considerably larger acreage among the *Brassica* crops. These crops are cultivated on an area of 6.51 million ha with a total production of 7.67 million tonnes, and with an average yield of 1179 kg/ha (Anonymous, 2011). It is cultivated in *rabi* season mainly in Northwest India, and contributes nearly 27 per cent to edible oil pool of the country (Singh *et al.*, 2010). Brassicas play an important role in the world agriculture as oilseed, vegetable, forage and green manure crops and condiments. Genus Brassica of family Brassicaceae (syn. Cruciferae), exists in a vast diversity of crop forms, unparalleled by any other genus in the family. Among the oilseed *Brassica*'s cultivated in our country, Indian mustard is considered to be the most important.

India is the fourth largest oilseed economy in the world. Among the seven oilseeds cultivated in India, rapeseed-mustard contributes 28.6% in the total oilseeds production and ranks second after groundnut sharing 27.8% in the India's oilseed economy. Inclusion of more diverse parents in hybridization programme increases the chances of obtaining maximum heterosis and gives a broad spectrum of variability in segregating generations. Keeping in view the importance of the subject matter, an attempt has been made to carry out the genetic diversity analysis in Indian mustard pertaining to Haryana state. Several researchers are working on similar lines regarding selection of diverse genotypes for breeding purpose at National/International level. Just to cite a few references in this regard; Srivastav *et al.* (2000), Acharya and Swain (2003), Muhammad *et al.* (2007), Misra and Kumar (2009) etc. have worked on genetic divergence in Indian mustard. Similar work on *Brassica spp.* at International level may be referred due to Zahan et al. (2008), Zaman *et al.* (2010), Shathi *et al.* (2012), Zada (2013) etc.

MULTIVARIATE STATISTICAL TECHNIQUES FOR SELECTION OF IMPORTANT CHARACTERS

Principal component analysis being a data reduction technique for investigating the interdependence attempts to simplify complex and diverse relationships existing among a set of observed variables, by revealing common dimensions or components that link seemingly unrelated variables. The procedure consists of finding the eigen roots and eigen vectors of the correlation matrix of explanatory variables. Interpretation of principal components is often facilitated by computing the components loadings. PC loadings are correlation coefficients between the PC scores and the original variables. PC loadings measure the importance of each variable in accounting for the variability in the PC. One of the most commonly used criteria for solving the number of components problem is the eigen value-one, also known as the Kaiser's (1960) criterion.

Cluster analysis is also one of the methods of data reduction technique. PCA reduces the number of variables whereas cluster analysis reduces the number of observations. Cluster analysis has similarity with discriminant analysis in respect of classification of observations. But discriminant analysis derives a rule for allocating an object to its known proper population based on some prior information of the group membership of the object, whereas the cluster analysis identifies homogeneous groups or clusters. It helps in grouping the materials in such a manner that similar types are grouped together while dissimilar ones belong to different groups. There are two main types of measures used to estimate this relation; distance measures and similarity measures. Hierarchical Cluster Analysis is a commonly used method for forming clusters and displaying similarities and dissimilarities between pairs of genotypes of a set by using one of the two methods; agglomerative or divisive. There is no unified approach on what actually constitutes a cluster. One of the important aspects of clustering is to study the differences among the formed clusters. Moreover, it is needed to decide the

appropriate number of clusters. One way MANOVA may be performed to check the accuracy of the clustering and to decide about the number of clusters.

The statistical technique of discriminant function is used to discriminate between/among various groups of objects when the dependent variable is categorical and independent variables are metric. The objective of discriminant analysis is to classify the sample objects accurately on the basis of a linear combination of predictor variables. Three different methods namely i) Maximum Likelihood Discriminant Rule, ii) Fisher's(1936) Linear Discriminant Function are iii) Bayes Discriminant Rule are in common use for identification of the populations. However, for the present study, the Fisher's linear discriminant function method was used for the purpose.

EXPERIMENTAL DETAILS AND STATISTICAL ANALYSIS

The experimental material comprising of sixty genotypes of Indian mustard were planted during *rabi* season of 2011-12 at the research farm of Oilseeds section, Department of Genetics and Plant Breeding, CCS HAU, Hisar. Each genotype was grown in a plot size of 1.5m x 3m with a row to row of spacing 30cm and plant to plant distance as 10cm. All the recommended package of practices was followed to raise a good crop. Observations were recorded on five competitive individual plants excluding border plants in each genotype for the quantitative traits viz., number of lobes per leaf , leaf length (cm), leaf width (cm), days to 50% flowering, days to maturity, plant height (cm), primary branches per plant, secondary branches per plant, main shoot length (cm), number of siliquae on main shoot , siliqua density on main shoot, siliqua length (cm), number of seeds per siliqua, 1000- seed weight (g), seed yield per plant (g) and oil content (%).

The genetic variability parameters including mean, range, standard deviation and coefficient of variation (CV %) of various traits are given in Table 1. Karl Pearson correlation coefficients shown in Table 2 were obtained to see the association among the traits under consideration.

Table 1: Descriptive Statistics of Yield and Yield Contributing Traits

Variables	Mean	Range	Std. Deviation	CV (%)
No. of lobes/ leaf	6.60	5.00	1.04	15.8
Leaf length (cm)	37.27	27.40	5.55	14.9
Leaf width (cm)	13.79	11.60	2.70	19.6
Days to 50 % flowering	51.95	25.00	7.14	13.8
Days to maturity	154.13	21.00	4.60	3.0
Plant height (cm)	209.10	103.00	23.71	11.3
Primary branches/ plant	6.61	8.00	1.61	24.4
Secondary branches/ plant	12.61	10.00	2.09	16.6
Main shoot length (cm)	60.30	35.00	8.38	13.9
No. of siliquae on main shoot	41.06	55.00	9.77	23.8
Siliqua density on main shoot	1.52	2.26	0.34	22.8
Siliqua length (cm)	4.27	30.40	3.81	89.3
No. of seeds/ siliqua	11.26	9.00	2.11	18.7
1000-seedwt (g)	4.25	27.37	3.43	80.6
Seed yield/ plant (g)	17.21	22.00	6.22	36.2
Oil content (%)	36.71	8.40	2.30	6.3
Valid N		60		

Table 2 : Pearson's Corrélation Coefficients among Quantitative Traits

Variables	No. of Lobes/Leaf	Leaf Length (cm)	Leaf Width (cm)	Days to 50% Flowering	Days to Maturity	Plant Height (cm)	Primary Branches / plant	Secondary Branches / Plant	Main Shoot Length (cm)	Siliquae on Main Shoot	Siliquadensity on Main Shoot	Siliqua Length (cm)	No. Of Seeds/ Siliqua	1000-Seed wt (g)	seedyield/ Plant (g)	Oil Content (%)
No. of lobes	1	.350**	.199	.090	.128	.073	-.032	-.025	.159	-.052	.223	.118	-.112	-.192	.204	.324*
Leaf length		1	.806**	.337**	.257*	.229	-.187	-.153	.201	.232	-.128	.001	.054	-.161	.115	.193
Leaf width			1	.330**	.250	.194	-.093	-.007	.148	.156	-.062	-.055	.025	-.104	.146	.115
Days to 50% flowering				1	.603**	.495**	.007	.139	-.070	.195	-.193	-.029	.109	-.115	.000	-.002
Days to maturity					1	.402**	-.032	.169	.006	.022	-.032	.137	.103	-.078	.025	-.157
Plant height						1	.157	.057	.197	.145	-.101	-.008	.071	-.167	.217	.002
Primary branches							1	.196	-.140	.197	-.213	-.147	.145	-.268*	.448**	-.050
Secondary branches								1	-.191	-.068	-.027	.013	.016	.106	.051	-.057
Main shoot length									1	.191	.380**	.034	-.055	.044	.080	.105
Siliquae on main shoot										1	-.743**	.069	.119	-.064	.266*	.165
Siliquadensity on main shoot											1	-.064	-.180	.090	-.147	-.052
Siliqua length												1	-.005	-.015	-.178	-.081
No.of seeds/ siliqua													1	-.006	.308*	.082
1000-seed wt														1	-.069	.030
Seed yield/ plant															1	.299*
Oil content																1

**. Correlation is significant at 0.01 level (2-tailed) , *. Correlation is significant at 0.05 level (2-tailed)

**. Correlation is significant at 0. 01 level (2-tailed), *. Correlation is significant at 0.05 levels (2-tailed)

Principal Component Analysis

The first seven PCs reflected about 74 % of the total variation giving an idea of the structure underlying the variables analyzed and indicating that the traits associated with these PCs are more useful in differentiating accessions (Table 3). The first PC showed higher loadings for the characters leaf length and leaf width. The second PC displayed higher loadings for days to maturity, days to 50% flowering and plant height. The higher loading displaying variables on 3rd PC were siliqua density and no. of siliquae on main shoot. The variables i.e., seed yield per plant/oil content, main shoot length, 1000-seed wt. and siliqua length were highly loaded on 4th, 5th, 6th and 7th PCs respectively (Table 4).

Table 3: Total Variance Explained by Different PC Components

Components	Eigen Values	% Variance	%Cumulative Variance
1	3.01	18.84	18.84
2	2.14	13.39	32.23
3	1.80	11.26	43.49
4	1.49	9.32	52.82
5	1.16	7.29	60.11
6	1.13	7.06	67.18
7	1.08	6.80	73.98
8	0.94	5.89	79.87
9	0.83	5.21	85.09
10	0.59	3.68	88.77
11	0.53	3.33	92.11
12	0.40	2.50	94.61
13	0.38	2.38	97.00
14	0.27	1.74	98.74
15	0.14	0.89	99.63
16	0.05	0.36	100.00

Table 4: Rotated Component Matrix Displaying Principal Variables

	Component(s)						
Variable(s)	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Leaf length (cm)	.881	.212	.123	.011	.148	-.065	.017
Leaf width(cm)	.822	.252	.059	.021	.042	-.021	-.134
Days to maturity	.150	.816	-.045	-.055	-.107	.019	.152
Days to 50% flowering	.310	.757	.144	-.019	-.153	-.012	-.024
Plant height (cm)	.036	.721	.060	.200	.265	-.169	-.041
Siliqua density on main shoot	-.063	-.032	-.916	-.072	.276	.084	-.014
No. of siliquae on main shoot	.122	.064	.847	.235	.219	-.039	.109
Seed yield/ plant (g)	.055	.051	.102	.838	.070	-.147	-.111
Oil content (%)	.402	-.313	.016	.558	.029	.150	.217
No. of seeds/ siliquae	-.067	.177	.244	.430	-.059	.204	-.125
Main shoot length (cm)	.068	.110	-.161	.143	.872	.125	.101
Secondary branches/plant	-.177	.308	-.159	.246	-.581	.176	.090
1000- seed wt. (g)	-.114	-.079	-.042	.038	-.012	.874	-.047
Primary branches/ plant	-.342	.107	.158	.505	-.158	-.526	-.206
Siliqua length (cm)	-.140	.103	.156	-.205	.033	.034	.846
No. of lobes/leaf	.448	-.022	-.377	.319	.030	-.276	.514
7 components extracted : Rotation Method: Varimax with Kaiser Normalization and converged in 15 iterations							

CLUSTERING AND DISCRIMINANT ANALYSIS

The mean, range, standard deviation and coefficient of variation (CV %) of yield and yield contributing traits showed that a considerable diversity exists in the experimental material. So, the data recorded on all these traits were subjected to cluster analysis and 60 genotypes were grouped into five clusters where each genotype within a cluster was closest to the cluster mean. The Ward's minimum variance method (1963) was used to carry out the agglomerative hierarchical cluster analysis. Thus, the cluster analysis helped in grouping the genotypes in such a manner that similar types are grouped together while dissimilar ones belong to different groups.

Further, the discriminant analysis was carried out for the selection of discriminator variables leading to the development of discriminant functions which were then used for classifying the observations. First three canonical discriminant functions were used for the purpose as is mentioned in Table 5. Tests for differences between the groups, considering all the variables simultaneously were dealt using Wilk's test statistic given in Table 6. To determine the inter-cluster distances, the data were analyzed on the basis of D^2-statistic to measure the genetic divergence among the genotypes and their average inter-cluster distances are shown in Table 8. However, the final cluster means in respect of all characters are given in Table 7. As there are genotypes superior for individual trait belonging to different clusters which indicates that none of the clusters contained genotypes with all the desirable characters. Thus, the genotypes superior for specific characters from different clusters may be selected for further utilization in breeding programme. Based on the

relative contributions of different characters; plant height, no. of siliquae on main shoot and days to maturity contributed more towards the genetic divergence and 98.3 % of the originally grouped cases were correctly classified as has been depicted in Table 9. Inter and intra-cluster distances from the group centroid are shown in Figure1 while the clustering pattern with name and number of genotypes in each cluster is expressed in Table 10.

Table 5: Percent Variance Explained by Discriminant Functions

Function(s)	Eigen Value	% Variance	% Cumulative Variance
1	6.74	79.4	79.4
2	1.70	20.0	99.4
3	0.05	0.6	100.0

First 3 canonical discriminant functions were used in the analysis

Table 6: Variables Entered /Removed

Step	Number of Variables	Wilks' Lambda	df1	df2	df3
1	Plant height	.131	1	4	55
2	Plant height No. of siliquae on main shoot	.062	2	4	55
3	Plant height No. of siliquae on main shoot Days to maturity	.046	3	4	55

At each step, the variable that maximized the Mahalanobis distance between the two closest groups was entered

Table 7: Final Cluster Means

Variables	Cluster(s)				
	1	2	3	4	5
No. of lobes/ leaf	6.83	6.45	7.00	6.40	6.64
Leaf length (cm)	37.36	34.60	36.60	37.25	42.08
Leaf width (cm)	13.49	12.86	14.60	14.10	15.61
Days to 50 % flowering	50.39	48.30	52.00	56.70	56.82
Days to maturity	153.06	152.05	154.00	156.40	157.64
Plant height (cm)	211.72	186.10	157.00	247.60	216.36
Primary branches/ plant	6.78	6.40	9.00	7.60	5.64
Secondary branches/ plant	12.17	13.20	12.00	13.30	11.73
Main shoot length (cm)	57.83	59.15	47.00	62.30	65.82
No. of siliquae on main shoot	36.94	36.60	80.00	45.30	48.55
Siliqua density on main shoot	1.60	1.65	.58	1.38	1.40
Siliqua length (cm)	3.90	3.74	3.70	3.64	6.47
No. of seeds/ siliqua	11.67	10.80	10.00	11.10	11.73
1000-seed wt (g)	4.27	4.81	4.75	3.48	3.89
Seed yield/ plant (g)	19.39	14.50	23.00	19.30	16.18
Oil content (%)	37.57	36.09	38.90	36.57	36.38

Table 8: Distances between Final Cluster Means

Cluster	1	2	3	4	5
1		26.46	70.67	37.85	18.14
2			54.71	63.19	35.85
3				98.50	70.79
4					32.41
5					

Table 9: Classification Results

		Cluster Number of Case	Predicted Group Membership					Total
			1	2	3	4	5	
Original	Count	1	18	0	0	0	0	18
		2	0	20	0	0	0	20
		3	0	0	1	0	0	1
		4	0	0	0	10	0	10
		5	1	0	0	0	10	11
	%	1	100.0	.0	.0	.0	.0	100.0
		2	.0	100.0	.0	.0	.0	100.0
		3	.0	.0	100.0	.0	.0	100.0
		4	.0	.0	.0	100.0	.0	100.0
		5	9.1	.0	.0	.0	90.9	100.0
98.3% of original grouped cases correctly classified.								

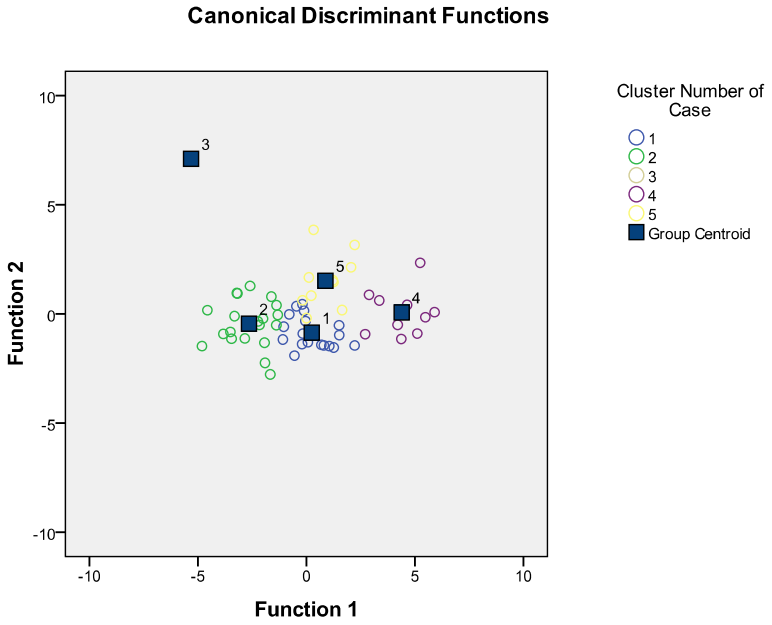

Figure 1: Inter and Intra-Cluster Distances

Table 10: Distributing Pattern of 60 Genotypes of Indian Mustard into Five Clusters

Cluster group	No. of Genotypes	Name of Genotypes
1	18	Varuna Albino, RC-1425, RH-9617, Sarita, Kranti, JMM937, JMMWR-9348, Pusa Bold, KM-888, RH-7846, RH0401(YS), RH-0406, RC-2, RC-5, RC-13 , RC-29 , RC-32 , RC-35
2	20	Parkash, RH-0345, BIO-902, Pusa Bahar, Shiva, RH-8701, RC-7, RC-14, RC-15, RC-20, RC-21, RC-22, RC-23, RC-24, RC-25, RC-27, RC-30, RC-31, RC-33, RC-34
3	1	RH0502
4	10	PM, RWH1, RC-199, RAURD25, Pahari rai , ZEM-2, RH- 8814, RC-6, RC-12, RC-18
5	11	RC-781, UDN-69, T-6342, RH-8912, EC126743, EC126745, ZEM-1, RH-0749, RC-8 , RC-26, RC-28

The statistical analysis showed that the sufficient variability exists in the material and cluster/discriminant analysis clearly helped in differentiating genotypes into major groups for various traits and to be used further for breeding purpose. The mean performance of different clusters calculated for different traits revealed wide range of differences among clusters with respect to these traits. The maximum inter-cluster distance was observed between clusters 3 and 4 (98.5) followed by clusters 3 and 5 (70.79), clusters 1 and 3 (70.67) etc. while the lowest inter-cluster distance was observed between clusters 1 and 5 (18.14) followed by clusters 1 and 2 (26.46), clusters 4 and 5 (32.41) etc. The genotypes from the clusters showing higher inter-cluster difference could be utilized in the hybridization programme as crossing between diverse parents is likely to produce wide genetic variability among the progenies of the segregating generations.

REFERENCES

1. Acharya, N.N. and Swain, D. 2003. Hybrid performance in relation to genetic divergence in Indian mustard [*Brassica juncea* (L.) Czern & Coss].*Crop Research.* 25(2), 312-315.
2. Anonymous. 2011. Directorate of Economics and Statistics, Department of Agriculture and Cooperation. 4th Advance Estimates released as on 19.07.2011.
3. Anwinithul, Patil., Shanti., Charjan. S.U., Thakare, P.G. and Wankhade, M. 2004. Genetic divergence studies in Indian mustard. *Soil and Crops*, 14(2), 297-304.
4. Fisher, R.A. (1936). The use of multiple measurements in taxonomic problem. Ann. Eugen, London, **7**, 179-88.
5. Johnson, R. A. and Wichern, D.W. (2006). Applied Multivariate Statistical Analysis. Pearson Education.
6. Jolliffe, I.T. (1972). Discarding variables in a principal component analysis. I: Artificial data, applied statistics, 21, 160-73.
7. Kaiser, H. F. (1960). The application of electronic computers to factor analysis, Educational and Psychological Measurement, 20, 141-51.
8. Mahalanobis, P.C. (1936). On the generalised distance in statistics. Journal of Genetics, 41, 159-93.

9. Misra, A.K. and Kumar, A. 2009. Characterization of Indian mustard *(Brassica juncea* L.) germplasm for economic traits. *Cruceferae Newsletter,* 28, 27-30.
10. Muhammad, Akbar. Usman, Saleem., Tahira., Muhammad, Yaqub. And Nasim, Iqbal 2007. Utilization of genetic variability, correlation and path analysis for seed yield improvement in mustard *(Brassica juncea*). *Journal of Agricultural Research Lahore,* 45(1), 25-31
11. Singh. D, Arya, RK, Chandra, N, Niwas, R and Salisbury, P. 2010. Genetic diversity studies in relation to seed yield and its component traits in Indian mustard [*Brassica juncea* (L.) Czern & Coss.]. *J Oilseeds Brassica,* 1, 19-22.
12. Shathi, D. A., Arifuzzaman, M., Biswas, B. K., Hasanuzzaman, M. and Azad, A. K. (2012). Genetic divergence in mustard (*Brassica* spp. L*). African J. Plant Science*. 6(8), 239-243.
13. Srivastav., Manoj, K. and Singh, R.P. 2000. Genetic divergence analysis in Indian mustard [*Brassica juncea* (L.) Czern & Coss]. *Crop Research*, 20(3), 555-557.
14. Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc., 58, 236-44.
15. Zada, Muhamad., Zakir, Nahida., Rabbani, M. Ashiq., Shinwari, Khan Zabta. 2013. Assessment of genetic diversity in Ethiopian mustard (*Brassica Carinata* A. Brun) germplasm using multivariate techniques. *Pakistan. Journal of Botany*. 45 (SI), 583-593.
16. Zahan, M. I., Bhuiyan, M. S. R. and Hossain, M. S. (2008). Genetic divergence in Oleiferous *Brassica* species. *J. Sher-e-Bangla Agric. Univ*. 2(1), 1-6.
17. Zaman, M. A., Khatun, M.T., Ullah, M.Z., Moniruzzamn, M. and Rahman, M.Z. 2010. Multivariate analysis of divergence in advanced lines of mustard (*Brassica spp*) *Bangladesh Journal of Plant Breeding Genet*., 23(2), 29-34